

VU Research Portal

Estimating unconstrained customer choice set demand: A case study on airline reservation data

Haensel, A.; Erdman, J.; Koole, G.M.

published in

Journal of Choice Modelling
2011

DOI (link to publisher)

[10.1016/s1755-5345\(13\)70043-5](https://doi.org/10.1016/s1755-5345(13)70043-5)

document version

Publisher's PDF, also known as Version of record

[Link to publication in VU Research Portal](#)

citation for published version (APA)

Haensel, A., Erdman, J., & Koole, G. M. (2011). Estimating unconstrained customer choice set demand: A case study on airline reservation data. *Journal of Choice Modelling*, 4(3), 75-87. [https://doi.org/10.1016/s1755-5345\(13\)70043-5](https://doi.org/10.1016/s1755-5345(13)70043-5)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:

vuresearchportal.ub@vu.nl

Estimating unconstrained customer choice set demand: A case study on airline reservation data

Alwin Haensel^{1,*} Ger Koole^{1,†} Jeroen Erdman^{2,‡}

¹ Department of Mathematics, VU University Amsterdam, De Boelelaan 1081a, 1081 HV Amsterdam, The Netherlands

² transavia.com, Piet Guilonardweg 15, Gebouw TransPort, 1117 EE Schiphol Airport, The Netherlands

Received 15 February 2011, revised version received 20 September 2011, accepted 21 November 2011

Abstract

A good demand forecast should be at the heart of every Revenue Management model. Yet most demand models do not incorporate customer choice behavior under offered alternatives. We are using the ideas of customer choice sets to model the customer's buying behavior. The demand estimation method, as described in [Haensel and Koole \(2011\)](#), is based on maximum likelihood and the expectation maximization (EM) algorithm. The main focus of the paper is the application case on real airline reservation data. The reservation data, consisting of the airline's daily flight offers, is used to unconstrain the underlying customer demand in terms of price sensitivity. Using this demand information per choice sets, the revenue manager obtains a clear view of the real underlying demand.

Keywords: customer choice behavior; demand estimation and unconstraining; revenue management

1 Introduction

Accurate demand information is essential for the success of all kinds of sophisticated booking or pricing controls, short Revenue Management (RM). A non-scientific introduction to RM with its main concepts, tactics and execution steps is given by [Cross \(1997\)](#). Any successful RM systems needs customer information

*T: +31 (0)20 5987687, alwin.haensel@vu.nl

†T: +31 (0)6 24922626, ger.koole@vu.nl

‡T: +31 (0)20 6046376, jeroen.erdman@transavia.com



on the micro-market level. The information should not only contain the number of customers to expect, but also comprise information on customer behavior such as price sensitivity. Historical data mostly consists of sales information per price class or product. But customers who book the same price classes or buy the same products are not necessarily equal; possibly some of them are also interested in other products or would also buy the product at a higher price. As [van Ryzin \(2005\)](#) formulates on page 206: what is needed in RM research is a change from product demand models to models of customer behavior. Our focus is to estimate unconstrained demand per choice behavior from given sales data. Unconstraining methods, such as the ones compared in [Queenan et al. \(2007\)](#), are applied to estimate the true demand quantities in cases of stock-outs. We understand unconstraining in a broader sense: We are not only interested in the number of unobserved customers, but also in their specific choice behavior. Therefore, we apply the unconstrained demand estimation method described by [Haensel and Koole \(2011\)](#) to a dataset of real airline reservation data.

The article continues in the following section with the explanation of the airline dataset, followed by the choice set model in Section 3. The estimation results are presented in Section 4, and our general findings are concluded in the final section.

2 Airline Data

In our case study, we are able to work on real airline booking data of two routes provided by Transavia, which we will from this point simply call Route 1 and Route 2. Both routes are connecting the Amsterdam airport Schiphol (AMS) with a Spanish airport and there is only one competitor airline serving the same direct connection. Unfortunately, we have no information of historic competitor prices available for our analysis. The datasets consist of the booking information for 11 consecutive departure day of weeks, i.e., we fixed a certain weekday for each route and work with the data of 11 weeks. The separation of different weekdays is very common in the airline business and based on statistical test which show a higher dependency and more common characteristics for consecutive weekdays than for consecutive days. The total bookings per departure day and route are shown in Figure 1.

The usual possible booking horizon consist of several months and can span a period up to a whole year. Even so, we observe that most of the bookings are made in a much smaller time span, namely 12 weeks prior to departure. The average number of bookings per week are shown in Figure 2, where week 1 denotes the beginning of the booking horizon and week 12 the week including the departure day.

On both routes we have $F = 12$ fare classes which only separate in price, as given in Table 1.

There are no extra services or standards associated with different fare classes. Thus, the price is the only differentiator, so there is only one active fare class at

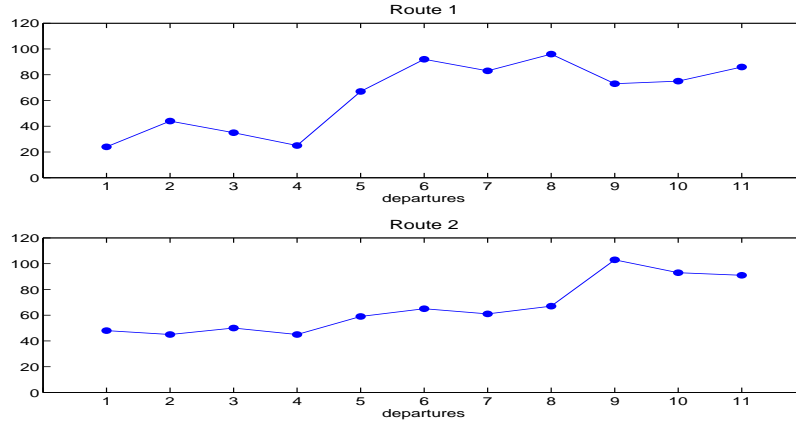


Figure 1: Total bookings for both routes and each departure.

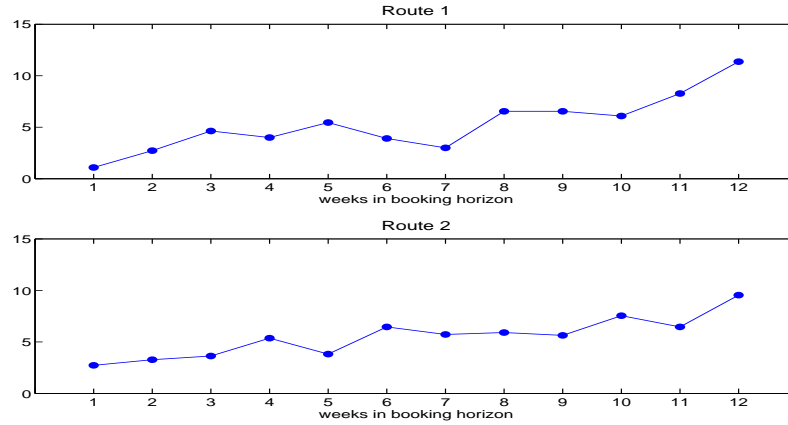


Figure 2: Average weekly bookings for both routes.

a time. The fare class booking and availability data is given on daily level. This means we know for each day in the booking horizon which fare class is available for booking, which is open, and also how many bookings are made. A whole flight can be unavailable for bookings if all fare classes are not available/closed. Table 2 shows the summarized information per fare class for both routes.

This information contains the number of departures when each fare class is open, the percentage on total booking days it is open ($11 \times 12 \times 7 = 924$ total booking days), the total number of sales/bookings and the averaged number of sales over all departures when the fare class was open. We find by adding the percentages of open days in Table 2 that Route 1 is 12.9% and Route 2 is 9.4 % of all considered booking days closed, i.e., no fare class is available. The case that all fare classes are closed can have two causes: First, there were no available seats

Table 1: Fare classes

most expensive -	Y	Z	S	B	M	H	Q	V	K	L	T	N	- cheapest
------------------	---	---	---	---	---	---	---	---	---	---	---	---	------------

Table 2: Performance data of Route 1 and 2

Route 1												
	Y	Z	S	B	M	H	Q	V	K	L	T	N
Dep. open	0	1	3	3	5	7	8	9	4	3	1	1
% open	-	0.6%	0.9%	3.4%	7.5%	20%	11%	29.5%	10.1%	3.7%	0.3%	0.1%
Total sales	0	2	18	24	48	149	105	160	45	45	7	7
Avg. sales	0	2	6	8	9.6	21.3	13.1	17.8	11.3	15	7	7

Route 2												
	Y	Z	S	B	M	H	Q	V	K	L	T	N
Dep. open	2	2	3	3	4	6	8	8	6	3	1	0
% open	0.2%	2.4%	4.3%	6.3%	6.6%	10.2%	16.6%	26.5%	15%	2.2%	0.5%	-
Total sales	2	17	39	40	101	76	136	148	121	26	12	0
Avg. sales	1	8.5	13	13.3	25.3	12.7	17	18.5	20.2	13	12	0

Dep. open - number of departures where this fare class is open, % open - fraction of possible booking days this class is open, Total sales - total sales per class over all departures, Avg. sales - averaged sale per class over open departures

and no further seats could have been sold. And second, when data points are removed by our outlier detection. From the datasets we have only information about the total sales per day, but not how they are made up. For example, if we observe six sales for a given day, we don't know if they are six individual sales or two sales of size 3, etc. We observe in our datasets some days with very large daily bookings, see Figure 3. The average daily booking size of Route 1 is 2.8 with a standard deviation of 2.2, for Route 2 we observe small values with an average of 2.5 and a standard deviation of 1.8. The extreme booking sizes are likely generated by group bookings, which we choose to exclude from our computation. Group bookings are normally not made via the usual online sales channels, but by direct negotiation with airlines representatives. Therefore, we will exclude booking days with more than seven bookings and the availability for these days is set to zero. Thereby, we are not overwriting sales data, we only exclude outlier data points from our estimation analysis.

3 Customer Choice Set Model

The concept of choice sets is earlier explained in [Ben-Akiva and Lerman \(1985\)](#). The proposed approach of unconstraining demand rate function per customer choice sets is described by [Haensel and Koole \(2011\)](#). Choice sets are sets of choice alternatives/product offers with a strict preference order. The customer's choice behavior is supposed to be represented by a choice set. The usual approach in travel demand modelling is to divide customers into different groups or segments based on their characteristics, best example is a segmentation in busi-

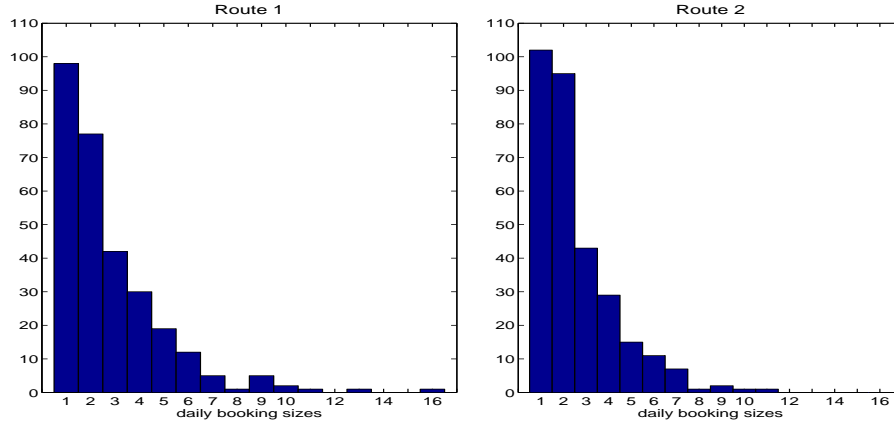


Figure 3: Histogram of daily booking sizes for both routes.

Table 3: Choice set example

Choice set c	$\{A\}$	$\{B\}$	$\{A, B\}$	$\{B, A\}$
Expected demand D_c	20	7	15	10

ness and leisure customers. Both segments are assumed to have very different buying behavior. The first are seen to buy more on short notice and are considered to have a high willingness to pay, and the second are supposed to be more price sensitive but book long in advance. The demand is then usually forecasted independently for each segment. In fact, our airline observes many leisure customers who have a relatively high willingness to pay and book close to departure, and on the other hand also observes early booking and price sensitive business customers. Therefore, our choice set approach is to distinguish customers by choice behavior independent of their individual characteristics. This can lead to choice sets made up by a very homogeneous customer group, but also allows a mix of different types of customers if their observed buying behavior is similar. The proposed demand estimation method associates demand quantities with different choice sets, representing different choice behaviors. Let us illustrate the choice set concept on a small example, where we consider an airline which offers two fare classes A and B . Fare A is the discount ticket consisting of the seat and no extra services and fare class B is the full ticket including a meal served during the flight. The possible choice sets are: $\{A\}$, $\{B\}$, $\{A, B\}$ and $\{B, A\}$, C will denote the set of all choice sets. Choice sets are written with a decreasing preference from left to right. Therefore, the choice set $\{A, B\}$ states that customers being represented by this choice set are strictly preferring ticket A over ticket B . In contrast, customers with choice set $\{B, A\}$ prefer the B over A . See Table 3 for some example choice set demand rates.

The airline can control the booking availability of all fare classes. \mathcal{O} denotes

the set of available products/open fare classes. The demand $D(f|\mathcal{O})$ of product/fare class f under offer set \mathcal{O} is defined by

$$D(f|\mathcal{O}) = \sum_{c \in C} D_c \cdot I_{U(c, \mathcal{O})=f}, \quad (1)$$

with I denoting the indicator function and $U(c, \mathcal{O})$ returns the fare class contained in choice set c with the highest preference under the set of offered alternatives \mathcal{O} or zero if $c \cap \mathcal{O} = \emptyset$. The amount of rejected customers, i.e., customers whose choice set c is non-overlapping with \mathcal{O} and therefore turned down, can be calculated by

$$D(0|\mathcal{O}) = \sum_{c \in C} D_c \cdot I_{U(c, \mathcal{O})=0}. \quad (2)$$

Hence, the sales probability of product $f \in \mathcal{O}$ is given by

$$P(f|\mathcal{O}) = \frac{D(f|\mathcal{O})}{D(0|\mathcal{O}) + \sum_{h \in \mathcal{O}} D(h|\mathcal{O})}. \quad (3)$$

The non-purchase probability is equivalently computed by

$$P(0|\mathcal{O}) = \frac{D(0|\mathcal{O})}{D(0|\mathcal{O}) + \sum_{h \in \mathcal{O}} D(h|\mathcal{O})}. \quad (4)$$

By definition, we set $P(x|\mathcal{O}) = 0$ if $x \notin \mathcal{O}$. Returning to our small airline example, the probability of selling a certain ticket to an arriving customer for different \mathcal{O} is

$$P(A|\mathcal{O} = \{A, B\}) = \frac{20 + 15}{0 + 20 + 15 + 7 + 10} = 0.67, \quad (5)$$

$$P(B|\mathcal{O} = \{A, B\}) = \frac{7 + 10}{0 + 20 + 15 + 7 + 10} = 0.33, \quad (6)$$

$$P(A|\mathcal{O} = \{A\}) = \frac{20 + 15 + 10}{7 + 20 + 15 + 10} = 0.87, \quad (7)$$

$$P(B|\mathcal{O} = \{B\}) = \frac{7 + 15 + 10}{20 + 7 + 15 + 10} = 0.62. \quad (8)$$

A more general example can be given by assuming the market to be segmented in such a way that different conditions are attached to the product itself, for example minimum length of stay, cancellation costs or membership credits. Each of these groups are again differentiated into different price categories. See Table 4 for an example airline portfolio.

Again, we assume that each customer has a set of classes which represent his willingness to buy (regarding price and conditions). These choice set may consist of any coherent sequence of subclasses, e.g., $\{L2, L1, M2\}$ or $\{M2, M1, H2\}$. The choice sets have again a strict preference order from left to right, i.e., customers whose behavior can be represented by choice set $\{M2, M1, H2\}$ want a possibility

Table 4: Choice set example 2, with groups: high, medium and low

Booking Class	Miles earned	Changes	Cancellations	Price
H1	100 %	charge 50	charge 100	520
H2	100 %	charge 50	charge 100	460
M1	50 %	charge 50	No	370
M2	50 %	charge 50	No	320
L1	25 %	No	No	250
L2	25 %	No	No	230

to change the ticket and their willingness to pay is larger or equal to 460 but strictly less to 520, since they are not willing to buy the H1 class. Of course the number of possible choice sets can be very high, but with some reflection of marketing and sales ideas we can usually restrict our selves to choice sets which are coherent and are of limited length. Another interesting possibility in choice set modelling is to incorporate competitor prices. To illustrate this, let us return to our first example with two fare classes, A and B . Say we have a competitor serving the same route and overing also two classes a and b , equivalently denoting the discount and full fare ticket. Our previous choice set $\{A, B\}$ is now divided into three corresponding choice sets which account for the different competitor pricing:

$$\{A, B\} \Rightarrow \left\{ \begin{array}{l} \{A[a, b, \emptyset], B[a, b, \emptyset]\} \\ \{A[a, b, \emptyset], B[b, \emptyset]\} \\ \{A[b, \emptyset], B[\emptyset]\} \end{array} \right\}$$

with $[\cdot]$ we denote the competitor's active classes under which we can observe a sale, \emptyset represents the case that the competitor does not offer any class. For example, $\{A[b, \emptyset], B[\emptyset]\}$ means that customers represented by this choice set prefer A over B for our company, but only buy class A if the competitor does not offer class a and they will only buy class B if we do not offer A and the competitor does not offer any class. For our dataset, we do not have any competitor pricing information available.

The example calculation (1)-(8) considers so far constant demands quantities. Usually the booking horizon is divided into T time stages $t = t_1, \dots, t_T$, where t_1 denotes the beginning of the booking horizon and t_T last time stage before the departure of the airplane. In general, we observe strictly increasing demand curves over the booking horizon, i.e., the demand increases towards the time of departure. Even though one often observes a drop in demand very close to the airplane's time of departure, the width of time stages can be defined such that the assumption of increasing demand over time stages is justified. The assumption can be relaxed to allow more complex demand functions, at the costs of additional parameters and the loss of structural properties of the estimation problem. In our

Table 5: Choice sets in airline test case

highest acceptable fare class	choice set
Y	N T L K V Q H M B S Z Y
Z	N T L K V Q H M B S Z
S	N T L K V Q H M B S
B	N T L K V Q H M B
M	N T L K V Q H M
H	N T L K V Q H
Q	N T L K V Q
V	N T L K V
K	N T L K
L	N T L
T	N T
N	N

case of European flights, we only observe a very small demand drop just before departure.

The estimation of the choice set model is divided into two steps: 1) The identification of different choice sets, and 2) The demand estimation per choice set. The identification of possible choice sets in our test case is very straightforward. The airline offers no extras with the airline seat, such as extra services or cancellation possibilities, and the considered flights are at most operated once day. Thus, the only product differentiation for a fixed itinerary is the price of the offered fare classes. Consequently the airline has only one available fare class at a time. All possible choice sets can therefore be given by combinations of consecutive fare classes (in fare order) starting with the cheapest class. Since price is the only differentiator, the cheapest fare class has the highest preference for all customers and the choice sets are only distinguishable by the upper willingness to pay. The choice sets in our airline test case are shown in Table 5.

For the second estimation step, the demand estimation per choice set, we apply the method proposed by [Haensel and Koole \(2011\)](#). They describe a parameter estimation method for the case of Poisson distributed demand with exponential demand functions $\lambda_c(t) = \beta_c \cdot \exp(\alpha_c \cdot t)$, for choice sets c and time stage t . The estimation method is based on maximum likelihood estimation (MLE) with an application of the EM-Algorithm. The open and closing decision of fare classes are on a daily level, but the time stages generally cover multiple days. Hence, we have to define $\mathcal{O}_t = \bigcup_{d \in t} \mathcal{O}_t(d)$ as the union of all sets of open classes $\mathcal{O}_t(d)$ for each day d in time stage t . Further, we have to redefine the $U(c, \mathcal{O}_t)$ for the input of sets of open classes by

$$U(c, \mathcal{O}_t) = \bigcup_{d \in t} \begin{cases} \{U(c, \mathcal{O}_t(d))\} & , \text{ if } U(c, \mathcal{O}_t(d)) > 0 \\ \emptyset & , \text{ else.} \end{cases} \quad (9)$$

The general log-likelihood functions of the estimation problem is given by

$$\mathcal{L} = \sum_{c \in C} \sum_{t=t_1, \dots, t_T} \log P \left[X = \mathcal{S}(t, c) | X \sim \text{Poisson}(\lambda_c(t) \cdot I_{\{U(c, \mathcal{O}_t) \neq \emptyset\}}) \right], \quad (10)$$

where $\mathcal{S}(t, c)$ denotes the number of sales/observed demand in time stage t corresponding to choice set c . In our input data, we have only the information of sales per day and fare class $S(d, f)$ and not per choice sets. As in [Haensel and Koole \(2011\)](#), we propose to use the EM algorithm, introduced by [Dempster et al. \(1977\)](#), to overcome this information problem in the MLE. The EM algorithm is an iterative method, where parameters are computed under an expectation based on values from previous iterations. In our application, we compute the expected number of sales at time stage t corresponding to choice set c in iteration i by

$$\mathcal{S}^i(t, c) = \left\lceil \sum_{d \in t} \frac{\lambda_c^{i-1}(t)}{\lambda_{\text{overlap}}^{i-1}(c, t, d)} \cdot S(d, U(c, \mathcal{O}_t(d))) \right\rceil, \quad (11)$$

where $\lambda_c^j(t)$ denotes the demand rate of choice set c at time t in the j^{th} iteration of the EM algorithm, $\lceil \cdot \rceil$ denotes the ceiling operator and $\lambda_{\text{overlap}}^j(c, t, d)$ is defined by

$$\lambda_{\text{overlap}}^j(c, t, d) = \sum_{s \in C} \lambda_s^j(t) \cdot I_{\{U(s, \mathcal{O}_t(d)) > 0 \text{ and } U(s, \mathcal{O}_t(d)) = U(c, \mathcal{O}_t(d))\}}. \quad (12)$$

Another problem occurs when the demand of choice set c is not observable at all days in times stage t , i.e., c is overlapping with $\mathcal{O}_t(d)$ for some but not all $d \in t$. Days within a time stage do not have the same booking intensity, e.g., we observe different booking intensities for different weekdays. The estimated booking intensity $\pi_t(d)$ of day d in time stage t is computed over historic booking horizons and reflect the different weighting between days in the same time stage and $\sum_{d \in t} \pi_t(d) = 1$. The definition of $\mathcal{S}^i(t, c)$ is extended to incorporate the booking intensity with an application of the rule of proportion by

$$\mathcal{S}^i(t, c) = \left\lceil \frac{\sum_{d \in t} \frac{\lambda_c^{i-1}(t)}{\lambda_{\text{overlap}}^{i-1}(c, t, d)} \cdot S(d, U(c, \mathcal{O}_t(d)))}{\sum_{d \in t} \pi_t(d) \cdot I_{\{U(c, \mathcal{O}_t(d)) > 0\}}} \right\rceil. \quad (13)$$

The starting values of the EM algorithm are obtained by ignoring the intersection of choice sets

$$\mathcal{S}^0(t, c) = \left\lceil \frac{\sum_{d \in t} S(d, U(c, \mathcal{O}_t(d)))}{\sum_{d \in t} \pi_t(d) \cdot I_{\{U(c, \mathcal{O}_t(d)) > 0\}}} \right\rceil. \quad (14)$$

λ_c^0 with the corresponding α_c^0 and β_c^0 parameters are obtained by minimizing the negative log-likelihood function separately per choice set

$$\begin{aligned} (\alpha_c^0, \beta_c^0) &= \arg \min_{\alpha, \beta > 0} -\mathcal{L}_c^0 \\ &= \arg \min_{\alpha, \beta > 0} -\sum_{t=1}^T \log \begin{cases} P(X_t = S^0(t, c)) & , \text{ if } U(c, t) > 0 \\ 1 & , \text{ else,} \end{cases} \end{aligned} \quad (15)$$

where $X_t \sim \text{Poisson}(\lambda = \beta_c \cdot \exp(\alpha_c \cdot t))$. The EM algorithm can be separated in two steps: The E-step, an application of equation 13 in the negative log-likelihood function for each choice set. Second, the M-step, which consists of minimizing $-\mathcal{L}_c$ separately for all choice sets $c \in C$. The algorithm stops if one of the following criteria is satisfied:

- maximum number of iterations reached,
- no changes in α and β values between iterations.

In general, we observe $\mathcal{L}^{i+1} = \sum_{c \in C} \mathcal{L}_c^{i+1} \geq \sum_{c \in C} \mathcal{L}_c^i = \mathcal{L}^i$. This results in the optimal solution of the MLE. Occasionally we observe that the EM algorithm reaches a likelihood maximum within the iteration cycle and that the final results hold a lower likelihood. In such a case we do not use the final EM output, but rather the intermediate results with the maximum likelihood.

4 Estimation Results

The proposed demand estimation method is tested on real airline reservation data to verify: 1) that the choice set model approximates the underlying demand closely, and 2) the estimation method is applicable for practitioners. In our estimation example we will consider the weeks in the booking horizon as time stages and the booking intensities π are obtained from the previous year's data. The estimation error is simply defined as

$$\text{error} = \text{actual} - \text{estimate}.$$

The demand estimate for any open fare class f at every booking day and departure combination is simply computed by

$$D(f|\mathcal{O}) = \sum_{c \in C} \lambda_c \cdot I_{U(c, \mathcal{O})=f}, \quad (16)$$

where \mathcal{O} denotes the given set of open classes, in our case a singleton. The choice set demand is estimated for all 11 departures in both datasets. In the following, we will examine the choice set estimation errors from different perspectives, such as: total relative errors over all booking days and price classes, errors per time

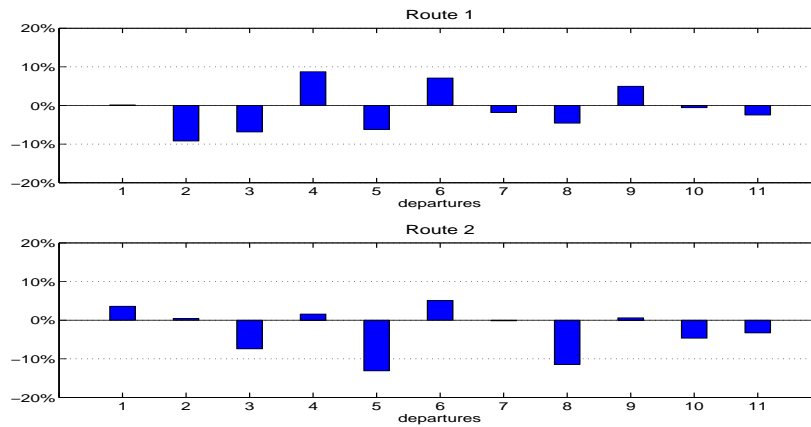


Figure 4: Total relative estimation errors per departure for both routes.

Table 6: Error Results per fare class for both routes

Route 1												
	Y	Z	S	B	M	H	Q	V	K	L	T	N
% open	-	0.6%	0.9%	3.4%	7.5%	20%	11%	29.5%	10.1%	3.7%	0.3%	0.1%
Avg. sales	0	2	6	8	9.6	21.3	13.1	17.8	11.3	15	7	7
Est. error	-	-0.5	2.3	-0.3	0.8	1.3	-0.5	-1.3	-1.0	-2.6	-0.1	-0.1

Route 2												
	Y	Z	S	B	M	H	Q	V	K	L	T	N
% open	0.2%	2.4%	4.3%	6.3%	6.6%	10.2%	16.6%	26.5%	15%	2.2%	0.5%	-
Avg. sales	1	8.5	13	13.3	25.3	12.7	17	18.5	20.2	13	12	0
Est. error	0.5	-1.8	-1.7	0	-1.5	1.4	-0.6	0	-2.3	1	0.6	-

% open - fraction of possible booking days this class is open, Avg. sales - averaged sale per class over open departures, Est. error - averaged estimation error per class

stages, and the error on fare class level. The total number of bookings over all fare classes and per departures is slightly overestimated by 1% for Route 1 and 2.6% for Route 2, i.e., the estimation error is negative if the estimate exceeds the actual. The total relative estimation errors for all departures are shown in Figure 4.

We observe no pattern of constant over or under estimation and please note that we are considering total numbers of usually less than 100 bookings. Figure 5 shows the average weekly bookings and the corresponding estimation errors. We observe, especially for Route 1, a slight constant overestimation in booking weeks 7-11. The last time stage, week 12, is underestimated for both routes. But when comparing the estimation errors to the average bookings, we find the errors to be considerably small in proportion.

Finally, we look into the estimation error on fare class level, with the results given in Table 6.

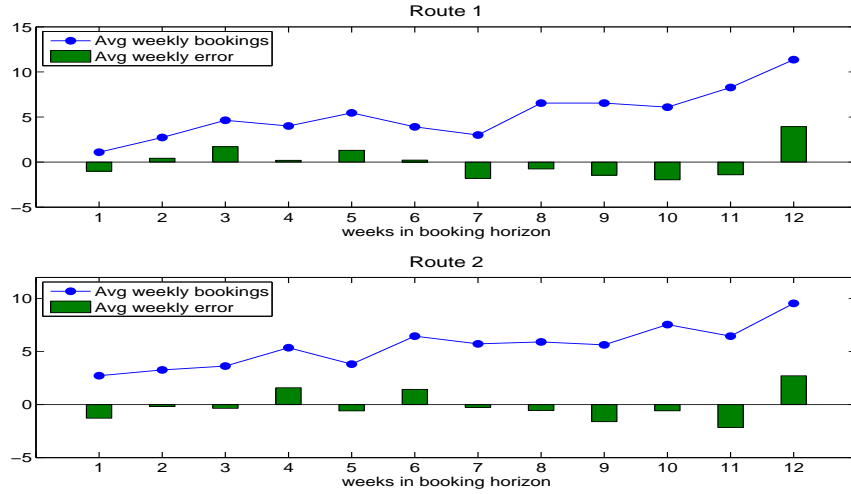


Figure 5: Average weekly bookings with the average approximation error for both routes.

As also shown in Table 2, we find that some fare classes are used much more often than other. Very high and very low classes are not often available for booking and thus we have limited data to estimate the corresponding choice sets. But even with this limited data, the estimation errors are considerably small compared to the average booking number when the considered fare class was available. The consideration of error results at more frequently used classes shows very low average errors for both routes. These results reinforce our positive conclusion on the proposed choice set based estimation method.

5 Conclusion

In this article, we study the problem of unconstraining sales data per price classes into demand estimates per customer choice behavior. Our proposed estimation method is tested on real airline data. The results show a slight overestimation on the total demand over all fare classes, but it should be noted that the total considered sales figures are usually smaller than 100. Hence, estimation errors of 5% are equivalent to an actual error of at most 5 bookings. Much more interesting than the results on the total bookings are the estimation errors on the fare class level. Here we observe small errors for all fare classes and especially very low values for frequently used classes. Further, the estimation method shows a very good computational behavior; the EM method converges in general within 10-15 iterations. This makes the algorithm feasible for practitioners. Overall, we find that the choice set model gives a very close approximation of the real underlying customer choice behavior, and that the estimation method can be successfully

implemented in real-world applications. Demand information on choice set level provides the revenue manager with detailed information on the price elasticity and the buying behavior of his customers. This information is crucial for any form of pricing or booking control. So far, the presented estimation method enables us to unconstrain choice set demand from given sales data. Further research on forecasting techniques at choice set level is needed to incorporate instant adjustments to demand fluctuations and shifts per choice sets.

References

- Ben-Akiva, M., Lerman, S., 1985. *Discrete Choice Analysis - Theory and Applications to Travel Demand*. MIT Press, Cambridge, MA.
- Cross, R., 1997. *Revenue Management: Hard-core tactics for market domination*. Broadway books, New York.
- Dempster, A., Laird, N., Rubin, D., 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society B* 39, 1–38.
- Haensel, A., Koole, G., 2011. Estimating unconstrained demand rate functions using customer choice-sets. *Journal of Revenue and Pricing Management* 10(5), 438–454.
- Queenan, C., Ferguson, M., Kapoor, R., Higbie, J., 2007. A comparison of unconstraining methods to improve revenue management systems. *Production and Operations Management* 16(6), 729–746.
- van Ryzin, G., 2005. Future of revenue management - models of demand. *Journal of Revenue and Pricing Management* 4(2), 204–210.